

# MtDNA sequence features associated with ‘selfish genomes’ predict tissue-specific segregation and reversion

Ellen C. Røyrvik<sup>1,2</sup> and Iain G. Johnston<sup>3,4,\*</sup>

<sup>1</sup>Department of Clinical Science, University of Bergen, Norway, <sup>2</sup>K.G. Jebsen Center for Autoimmune Diseases, University of Bergen, Norway, <sup>3</sup>Department of Mathematics, University of Bergen, Norway and <sup>4</sup>Alan Turing Institute, London, UK

Received May 15, 2020; Revised June 25, 2020; Editorial Decision July 12, 2020; Accepted July 15, 2020

## ABSTRACT

Mitochondrial DNA (mtDNA) encodes cellular machinery vital for cell and organism survival. Mutations, genetic manipulation, and gene therapies may produce cells where different types of mtDNA co-exist in admixed populations. In these admixtures, one mtDNA type is often observed to proliferate over another, with different types dominating in different tissues. This ‘segregation bias’ is a long-standing biological mystery that may pose challenges to modern mtDNA disease therapies, leading to substantial recent attention in biological and medical circles. Here, we show how an mtDNA sequence’s balance between replication and transcription, corresponding to molecular ‘selfishness’, in conjunction with cellular selection, can potentially modulate segregation bias. We combine a new replication-transcription-selection (RTS) model with a meta-analysis of existing data to show that this simple theory predicts complex tissue-specific patterns of segregation in mouse experiments, and reversion in human stem cells. We propose the stability of G-quadruplexes in the mtDNA control region, influencing the balance between transcription and replication primer formation, as a potential molecular mechanism governing this balance. Linking mtDNA sequence features, through this molecular mechanism, to cellular population dynamics, we use sequence data to obtain and verify the sequence-specific predictions from this hypothesis on segregation behaviour in mouse and human mtDNA.

## INTRODUCTION

Mitochondria are metabolically central organelles contained in almost all eukaryotic cells. Due to their evolution-

ary history, mitochondria retain genomes (mtDNA) which encode machinery important for mitochondrial functionality. There may be hundreds or thousands of copies of mtDNA in a given cell, forming a dynamic population of individual molecules which replicate and degrade quasi-independently of the cell cycle (1).

As a result of mutation, experimental manipulation, or gene therapies, mtDNA populations within cells may consist of a mixture of different mtDNA sequences co-existing. This situation is known as heteroplasmy (2,3). In heteroplasmic mixtures, segregation bias—the systematic proliferation of one haplotype over another over time—is commonly observed (1–2,4). For example, a cell starting with a mixture of 50% haplotype A and 50% haplotype B may after one year retain 80% haplotype B and only 20% haplotype A. This segregation is of pronounced medical importance as it can lead to the amplification of a disease-causing mutation if present on the proliferating haplotype (4). However, the causes of segregation bias remain largely unknown, challenging our basic biological understanding and our ability to optimally plan gene therapies.

Most systematic work on segregation bias has been performed in mouse models (5–13), where pairs of haplotypes in a heteroplasmic context are seen to exhibit complex but consistent tissue-specific segregation behaviour (Table 1). For example, segregation bias is often observed to favour one haplotype in liver and a different haplotype in blood (with this behaviour observed consistently across different hematopoietic cell populations (9)). Tissue-specific directions of segregation bias appear consistent across different nuclear backgrounds (8,9) (although the magnitudes of segregation bias may vary according to nuclear context, with several nuclear loci implicated (9,14–15)). Segregation bias has also been observed in *Drosophila* (16,17) and livestock (18,19) models, including recent work in minipigs (20), and with tunable levels of mtDNA selection in *Drosophila* (21). Recently, human stem cell assays have also shown segregation bias (22,23). Here, mirroring gene therapies, heteroplasmic cells are created using a maternal cell from which

\*To whom correspondence should be addressed. Tel: +47 555 82838; Email: iain.johnston@uib.no

most mitochondria are removed, and a ‘donor’ cell providing a replacement supply of mitochondria. In many maternal-donor pairings, the donor mtDNA remains stable, but some pairings exhibit ‘reversion’: segregation bias that favours the initially small amount of maternal mtDNA and amplifies the maternal type to dominate the cell. This reversion presents a potential issue for the recently UK-approved mitochondrial replacement therapies, where technical limitations entail the creation of heteroplasmy with a minority presence of pathogenic mtDNA types that could come to outcompete non-pathogenic molecules (4).

The basis for this segregation bias remains poorly understood, despite substantial recent discussion motivated by this importance for gene therapies (22,24–25). Some nuclear-encoded genes influence somatic segregation bias (9,14–15). Some evidence exists for segregation bias being more common and stronger in pairings of more genetically diverse haplotypes, suggesting that mtDNA sequence differences play at least some mechanistic role (12,13). However, further observations suggest that genetic distance alone is not sufficient to predict segregation bias (23). It is therefore likely that differences at specific regions play a role, and overall genetic distance just increases the probability of differences at these specific regions.

Fundamentally, the proliferation of a particular mtDNA type in a tissue depends both on the replication of individual mtDNA molecules and the survival of the organelles and cells that contain them. This survival relies on the expression of respiratory machinery from mtDNA. MtDNA has a unique ‘replication-transcription switch’ (26): unlike the situation in the nucleus, replication of mitochondrial DNA is wholly and directly dependent on transcription, by the single, standard mitochondrial RNA polymerase, POLRMT. POLRMT initiates transcription at the light strand promoter (LSP), and will either continue, producing a full polycistronic transcript, or be attenuated, forming an RNA primer that the replicative mitochondrial DNA polymerase (pol  $\gamma$ ) can use to produce another copy of the mitochondrial genome (27).

Many factors affect mtDNA replication and mitochondrial fitness and survival. Cell-wide factors including TFAM and POLRMT (28,29) and bioenergetic and metabolic demands affect the levels of replication and transcription throughout the whole mtDNA population. Mutations in the coding regions of mtDNA may lead to a subset of molecules within the cell producing dysfunctional or toxic gene products; subsequent quality control of mitochondria can clear these pathological mutations (30). Each of these whole-cell and sequence-specific factors can influence mtDNA dynamics and cell fitness.

Here, we consider a different class of mtDNA feature proposed to act in concert with these whole-cell and within-cell factors: sequence features that do not directly influence the functionality of gene products, but affect the balance of replication and transcription for a given mtDNA molecule. This balance of replication and transcription contributes to the *selfishness* of a given mitochondrial genome (31,32). Consider a cell with an admixture of two mtDNA types, with negligible differences in the gene products arising from these two types. If one type undergoes replication more readily, and one type more readily undergoes

transcription, the first type may be regarded as more ‘selfish’, favouring its own proliferation over the contribution of functional metabolic components to the cell. In this picture, we consider the functionality of gene products of different mtDNA molecules to be equivalent but not necessarily perfect. If one sequence harbours variants that reduce functionality or cause other cellular issues relative to the other, this constitutes an independent axis of selection that will shape mtDNA dynamics in addition to the replication-transcription balance (1).

Specific examples of selfish mtDNA have been known for some time (33,34), including in plants (where mtDNA features can cause male sterility (35,36)), yeast (where ‘petite’ mutants are highly replicative without encoding the usual complement of mitochondrial machinery (37)), and human (where a selfish deletion mutant causes myopathy (38)). Theoretical work has suggested that different levels of organelle- or cell-level selection can influence the balance of selfish and less selfish mtDNA types (36). Elegant experiments have accordingly shown that different balances between selfish drive and purifying selection result in different patterns of mtDNA proliferation in *Drosophila* (21) and yeast (37), and shown the interactions between ‘hitchhiking’ selfish mtDNA and cellular control in nematodes (39).

In this article, we show that this replication-transcription balance coupled with cell-level selection can account for complex observed tissue-dependent mtDNA segregation patterns. Synthesising published *in vitro* biochemical data, we identify specific sequence features, related to G-quadruplex formation, that predict the behaviour of an individual mtDNA haplotype under this model and show that these features predict both tissue-specific mtDNA behaviour in mice and haplotype-specific reversion in human stem cells. Taken together, our results suggest that the stabilities of G-quadruplex forming sections of the mtDNA control region determine sequence ‘selfishness’ and are important contributors to tissue-specific mtDNA segregation bias.

## MATERIALS AND METHODS

### Mouse data

We use observations from (5,7,8,11–13) and report those tissues where statistically significant segregation bias was reported. Where statistical calculations were absent, we report tissues where segregation bias relative to a reference tissue (typically lung, generally observed to show weak if any segregation bias) was consistently observed.

### Human stem cell data

The correction (40) to human mtDNA data in (22) was used to reconstruct human mtDNA sequences, which are given for reference in Supplementary Information. The labels in Figure 4 refer to specific pairings from (40), with haplotypes: (1) U5a/H1b; (2) U5a/H1b; (3) B2k/H49; (4) X2c/D4a; (5) H56/H2a; (6) H2a/H56; (7) H2a/H56; (8) H44a/H13a; (9) H1b/U5a; (10) H1b/U5a; (11) U5a/H1g; (12) H1g/U5a; (13) h1e/D1f; (14) h1e/D1f; (15) h1e/D1f; (16) D4a/A2g; (17) A2g/D4a; (18) T2b/T2; (19) B2k/H49;

**Table 1.** Segregation bias in mouse models. Nuclear context and haplotype pairs for which tissue-specific segregation bias is observed in mouse models, and the direction (or absence) of that bias. <sup>1</sup> Significance patterns of observed segregation are ambiguous due to presentation. <sup>2</sup> Segregation patterns are ambiguous due to absence of reference tissue, but liver and brain segregate significantly differently

Reference(s)	Nuclear context	mtDNA type 1	mtDNA type 2	Type 1 increase	Type 2 increase	Neither/ambiguous
5,9–10)	(various, diverse)	NZB	BALB	Liver, kidney	Blood, spleen	Brain, muscle, lung, tail, heart
(8)	RR × DBA or C57BL/6	RR	C57BL/6	Liver, kidney, ovary, stomach, gut (relative to lung <sup>1</sup> )	Brain, heart (relative to lung <sup>1</sup> )	Spleen? (relative to lung <sup>1</sup> )
(11)	JF1 × 129	JF1	C57BL/6	Liver? (relative to tail <sup>2</sup> )	Brain? (relative to tail <sup>2</sup> )	Kidney, tail <sup>2</sup> ?
(7)	C57BL/6	NZB	129S6	Liver, kidney, tail, brain	Skeletal muscle, seminal vesicle, ovary, pancreas, spleen	Lung, heart
(12)	C57BL/6	LE	C57BL/6	Gut, spleen, liver, kidney, lung	-	Testis, tail, skin, uterus, blood, heart, muscle, brain
(12)	C57BL/6	HB	C57BL/6	Blood, spleen, liver, lung	Heart, muscle	Gut, testis, tail, skin, uterus, kidney, brain
(12)	C57BL/6	BG	C57BL/6	Gut, testis, tail, skin, uterus, blood, spleen	-	Liver, kidney, lung, heart, muscle, brain
(12)	C57BL/6	ST	C57BL/6	Gut, testis, tail, skin, uterus, blood, spleen, liver, kidney, lung, muscle	-	Heart, brain
(13)	NZW	NZW	C57BL/6	-	-	Brain, heart, lung, liver, kidney, stomach, gut, spleen, muscle, adipose, skeleton, bladder, gonads, brain, skin

(20) H1b/H56; (21) H1b/H56; (22) H1b/H56; (23) H1b/H56; (24) F1a/D4a; (25) F1a/H1b; (26) X2c/U5a.

### Human haplotype data

The set of human mtDNA sequences available from NCBI was downloaded and custom code was used to extract the G-quadruplex regions in CSB2 and the TAS-proximal area and the haplotype information included in the accessions (directly available for 6630 records). The full list of accessions and structures is contained in Supplementary Information.

### Other animal data

Ref. (18) has a bovine *Bos indicus* nuclear background, *B. taurus* mtDNA and *B. indicus* mtDNA. The *B. taurus* type is increased through fetal development. Ref. (19) has a porcine Meishan nuclear background, Meishan mtDNA and Landrace mtDNA. Compared to ear, Meishan mtDNA was higher in liver and lower in spleen and blood, although the statistical significance of these effects is not clear.

Accessions used are as follows. *Mouse haplotypes*. JF1 (KR020498.1); NZB/B1NJ (L07095.1); C57BL/6J (AY172335.1); C57BL/6N (KR020497.1); BALB/cByJ (EF108333.1); RR D-loop from Ref. (8) (AB025348.1); C57BL/6 D-loop from Ref. (8) (AB033825.1); LE (KC663618.1); BG (KC663619.1); HB (KC663620.1);

ST (KC663621.1); NZW (EF108341.1); 129S6 (differences from L07095.1 listed in SI of (7)). *Pig haplotypes*. Meishan (KM998967.1); Landrace (NC\_000845.1). *Cattle haplotypes*. *Bos taurus* (NC\_006853.1); *Bos indicus* (NC\_005971.1). We performed all alignments with Clustal Omega (41). Alignments of cattle and pig are shown in Supplementary Figure S7; mouse and human stem cell alignments are available online (see Data and Code Availability).

### Modelling mtDNA agents and cell populations

Our modelling framework represents mitochondria as individual agents with a genetic identity and a complement of respiratory proteins, subjected (or not) to selective pressure based on these protein complements. We model a cell as containing  $m$  mtDNAs, each of which may be of type 1 (with selfishness  $\lambda_1$ ) or type 2 (with selfishness  $\lambda_2$ ). A parameter  $m^*$  dictates the maximum mtDNA content of a cell; we set  $m^* = 1000$  but verified that changes to this parameter did not strongly influence model behaviour. Each mtDNA  $i$  exists in a mitochondrial element, which also contains an amount  $p_i$  of respiratory protein complexes (we will later allow mixing of content between these elements to capture mitochondrial networks and dynamics, see below). Each mtDNA experiences polymerisation initiation events which are random (Poissonian) and occur with rate  $\gamma(1 - m/m^*)$ . An event for mitochondrion  $i$  leads to replication with probability  $\lambda_i$  (we will later reduce this probability to capture

the probability of replication failure, see below), and transcription with probability  $(1 - \lambda_i)$ . If mtDNA  $i$  replicates, another mtDNA of the same type is added to the population, and it and mtDNA  $i$  are both assigned  $p_i/2$  proteins. If mtDNA  $i$  undergoes transcription,  $p_i$  is increased by 1.

Each mtDNA experiences random protein degradation events with rate  $\kappa_p$ , where  $p_i$  is decreased by 1, and random potential mtDNA degradation events with rate  $\kappa_m$ . In this event, mtDNA  $i$  is removed from the population if  $p_i < P$ . If  $p_i \geq P$ , nothing happens for this event.  $P$  is thus a threshold amount of respiratory machinery required for adequate mitochondrial function.

We fix  $\lambda_1$  and consider the balance of haplotypes over time with different  $\lambda_2$  and  $P$ . We simulate the system for 1000 timesteps (enough for transient behaviour to disappear and a consistent readout of selective shifts to emerge, Supplementary Figure S2) for  $\gamma = \kappa_m = 0.1$ ,  $\kappa_p = 0.01$  (different values do not qualitatively change our results). Supplementary Figure S2 shows typical time behaviours of copy number, heteroplasmy and protein statistics over time in the simulation. Figure 2 uses  $\lambda_1 = 0.3$ ; consistent behaviour is observed for different  $\lambda_1$  (Supplementary Figure S1).

In the Supplementary Information we describe adaptations of this model to consider cell-level rather than organelle-level selection and show the same behaviour under several model structures and parameterisations. We also consider structural variations of the core model. First, we model the possibility of the failure of replication initiation with a parameter  $\rho$  describing the probability that, when the decision is made to replicate an mtDNA molecule, the replication event fails to initialise and so no change is made to the population (Supplementary Figure S3). This noise partially decouples effective replication rate from the current state of the system  $m$ . Second, we model the sharing of protein content between mitochondrial elements with a diffusion-like term (Supplementary Figure S4). Specifically,  $p_i = D(\langle p \rangle - p_i)$ , where  $\langle p \rangle$  is the mean protein content across all mitochondrial elements. High  $D$  homogenises the population;  $D = 0$  keeps the protein complements associated with different mtDNA molecules entirely separate.  $D > 0$  thus accounts for a ‘leaky’ genotype-phenotype relationship between an mtDNA molecule and its protein products, which may arise through exchange of mitochondrial content via mitochondrial dynamics and network structure (42,43).

### Statistical methods

For model comparison, we used Bayesian logistic regression with the response variable ‘did reversion occur?’ and predictor variables describing the differences in guanine counts in CSB2 and the three TAS regions. This Bayesian regression was implemented using the *arm* package in R (44), using an approximate EM algorithm and Student-t prior distributions for regression coefficients, as numerical issues occurred when fitting some model incarnations using built-in generalized linear model functionality. The Bayesian Information Criterion was used to select the best model; the corrected Akaike Information Criterion gave comparable results although suggested a role for TAS.G4a as well as TAS.G4c (Supplementary Table S1). Predictive

performance was calculated by holding back half the non-reverting and half the reverting observations as a test set, training the logistic model on the remaining data, then assessing its predictions on the test set. This procedure was performed  $10^3$  times with different samples of half the observations in each case. Analysis was performed in R (45); McFadden’s pseudo- $R^2$  value was computed using the *pscl* package (46).

## RESULTS

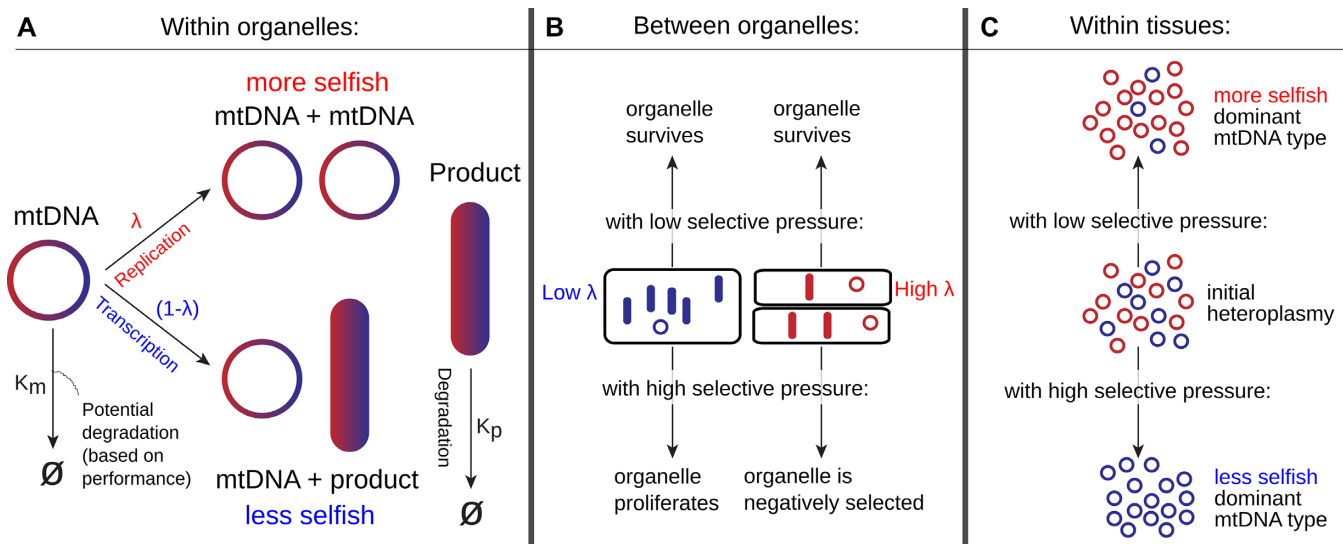
### A replication-transcription-selection model predicts saddle-like patterns of mtDNA segregation bias

We first sought to construct a general theory with which to understand and predict tissue-specific patterns of mtDNA segregation. Following the above idea of the replication-transcription switch (26), we assume that some sequence features of a given mtDNA molecule (to be determined later) determine that molecule’s balance between replication and transcription. MtDNA molecules favouring replication are in a sense more ‘selfish’, more readily undergoing replication and less readily producing respiratory machinery of use to the cell. MtDNA molecules favouring transcription are correspondingly less ‘selfish’. Over time, each molecule undergoes replication and transcription events, respectively producing more mtDNA (of no direct value to the cell) and polycistronic transcripts (which can be translated and contribute to the cell’s bioenergetic poise) according to its particular balance (Figure 1A).

We reiterate that many other factors will influence the transcriptional activity of a given mtDNA in a given tissue, including signalling and nuclear control (for example, via TFAM expression (28)), levels of polymerase (29), metabolic poise, and others. However, our theory focusses purely on the *relative* difference in activity between mtDNA types, and its predictions hold under variability in these overall levels (see below).

We then consider a situation where two mtDNA types, with different replication-transcription balances, co-exist in the same cell. We then have a more selfish mtDNA type favouring replication co-existing with a less selfish mtDNA type favouring transcription. In the absence of any other effects, the more selfish type will replicate more quickly, and over time come to genetically dominate over the less selfish type (Figure 1B and C).

However, the bioenergetic capacity of the mtDNA’s ‘host’ mitochondria must also be considered. Selfish mtDNA types will produce less respiratory machinery and contribute less to bioenergetic performance. Organelles containing selfish mtDNA types may then be expected to experience more functional challenges than those containing less selfish mtDNA types, which produce more machinery and contribute more to performance. If organelles are selected for bioenergetic capacity, as is known to occur through mitophagy and recycling (30), those mitochondria containing less selfish mtDNAs will then preferentially survive, while those less functional mitochondria containing more selfish mtDNA will be recycled. This selection acting at the organelle and/or cellular level can then balance, or reverse, a replicative advantage at the molecular level, as found in quantitative models of plant mtDNA populations (36) and



**Figure 1.** The replication-transcription-selection model predicts different segregation bias in different tissues. Individual mtDNA molecules may undergo selfish replication (with rate  $\lambda$ ) or co-operative transcription and gene expression (with rate  $(1 - \lambda)$ ). Cells containing many selfish mtDNAs will quickly accumulate more selfish mtDNA, but will lack respiratory machinery; cells containing less selfish mtDNAs will less quickly accumulate less selfish mtDNA, but will have higher levels of respiratory machinery. If there is selective pressure for respiratory capacity (most manifest in tissues with high cell turnover), cells dominated by selfish mtDNAs will be removed in favour of those with less selfish mtDNAs that have maintained cellular performance. Where there is little selective pressure (or little cell turnover to manifest such pressure), the selfish mtDNAs will come to dominate the tissue due to their unchecked replicative advantage.

in models and experiments on yeast (37) and *Drosophila* (21).

It is thus easy to see how selfishness can be either a successful or unsuccessful strategy depending on the level of selection for bioenergetic capacity that acts within cells (Figure 1B and C). In tissues with low mitochondrial and/or cellular turnover, and/or low energetic requirements, nothing prevents selfish mtDNAs proliferating and dominating the tissue. By contrast, when higher organelle and/or cellular turnover allows mitochondria to be selected based on bioenergetic performance, unselfish mtDNAs will come to dominate (as selfish mtDNAs compromise the performance of their host organelle).

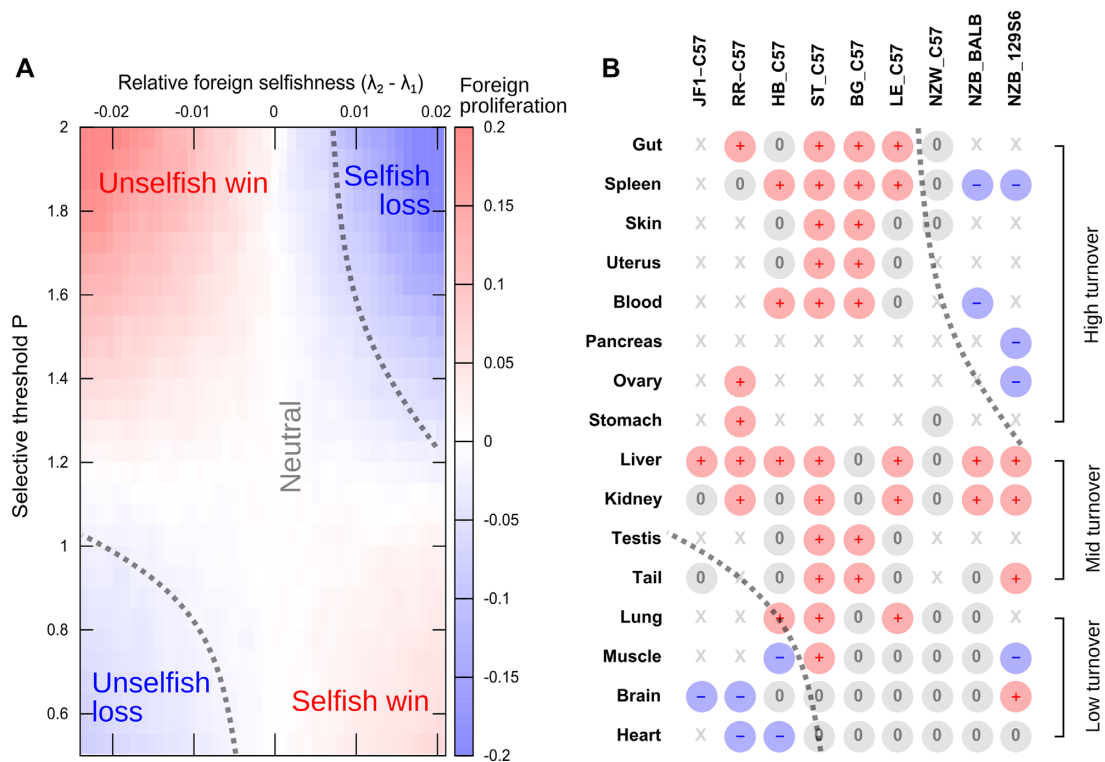
We underline that we are suggesting a mechanism that describes the *relative difference* in replication and transcription between two mtDNA types in the same cell. TFAM, POLRMT, bioenergetic and metabolic poise, and many other dynamic influences affect the *global* amount and balance of replication and transcription across mtDNAs (28,29). Our theory describes the *relative* difference in behaviour of two mtDNA sequences on this given global background. If, for example, global rates of mtDNA replication are limited, the relative replicative poise of each mtDNA type can remain intact, so that the overall patterns of relative proliferation remain unchanged. Further, we underline that our suggested mechanism acts *in addition* to selection arising from functional or other differences between mtDNA gene products. Here we assume that mtDNA sequences differ only in their replication-transcription balance, with no fitness differences due to gene product function or production. Any such differences will constitute another axis of organellar selection (see Discussion).

To translate this theoretical picture into a quantitative model, we assign each mtDNA type a parameter  $\lambda$  de-

scribing its replication-transcription balance, or, equivalently, its ‘selfishness’. Replication or transcriptional events take place at random, following the ‘relaxed replication’ paradigm (47) that has been employed in related recent studies on mtDNA dynamics (48). Each time an event occurs,  $\lambda$  gives the probability that the mtDNA molecule will replicate;  $(1 - \lambda)$  is correspondingly the probability that it will follow the transcription route.

Several previous studies have considered the dynamics of mixed mtDNA populations with functional differences between types (49). However, these studies do not typically consider expression of respiratory machinery (50–53), although it can be pictured as a general mitochondrial quality state (37,54). To explicitly capture the transcription-replication balance, we consider each mtDNA to be contained within a mitochondrial element. Transcription increases the amount of respiratory machinery in this element; replication induces a division into two elements, each inheriting half the previous element’s complement of respiratory machinery. Mitochondrial elements are randomly marked for degradation, and degradation proceeds if a marked element possesses less than a given threshold amount of respiratory machinery.

As shown in Figure 2A and Supplementary Figure S1, this simple model predicts that the magnitude and direction of segregation bias will depend on the selfishness difference between mtDNA types and the level of selection acting at the organelle level. Specifically, ‘saddle’-like behaviour is predicted, where an mtDNA type will be favoured if it is highly selfish in low-turnover tissue, or if it is less selfish in high-turnover tissues. When little selection acts, selfish mtDNA types can replicate without consequence and come to dominate the cell. By contrast, when selection for bioenergetic capacity is imposed, organelles (or cells) containing



**Figure 2.** The RTS model predicts the saddle-like tissue-dependent segregation bias observed in mouse models. (A) Segregation bias under the RTS model, as a function of the selfishness difference ( $\lambda_2 - \lambda_1$ ) between mtDNA types (horizontal axis) and the strength of organelle-level selection, manifest as the threshold value of respiratory protein content  $P$  required for organelle survival (vertical axis). At low levels of selection, the more selfish mtDNA type proliferates; at higher levels of selection, less selfish mtDNAs are favoured due to their improved contribution to bioenergetic performance. (B) Observed patterns of segregation in mouse mtDNA pairings from Table 1. Tissue-specific behaviour follows the predictions of the RTS model and the resulting ordering of haplotype pairs provides an estimate of relative selfishness for each pairing.

selfish mtDNAs are functionally challenged and therefore removed, leading to an overall accumulation of less selfish mtDNAs. The overall picture involves the direction of segregation bias changing (from favouring selfish to favouring less selfish mtDNAs) as selection increases, producing a saddle-like structure (Figure 2A).

This saddle-like behaviour is predicted robustly under different structures and parameterisations of the RTS model (Supplementary Figure S2). In particular, it is only the *relative* rate of transcription versus replication (given by  $\lambda$ ) for a given mtDNA molecule that influences the direction of segregation bias; the *absolute* rates just influence the timescale at which this bias becomes manifest. The structure of the saddle-like behaviour remained unchanged when we introduced noise in replication initiation (Supplementary Figure S3) and partial homogenisation of protein content between mitochondrial elements (Supplementary Figure S4), with the latter representing the case where protein content is shared through a mitochondrial network or through transient fusion events (42).

The results in Figure 2A show the mean behaviour of haplotypes, but deviations from this mean behaviour are to be expected due to intrinsic cellular noise. Under the RTS model, cell-to-cell variability in mtDNA proportions increases linearly with time, following predictions of (48) and observations of (55) (Supplementary Figure S2). We also consider a variant of this model where selection instead acts at the cellular level and show that the results are

qualitatively identical (Supplementary Figure S5). These general observations agree with previous studies investigating the action of organelle- or cell-level selection on differently-replicating molecules in the context of plant mtDNA (36).

### The RTS model accounts for the complex patterns of tissue-specific segregation in mouse models

We next asked whether our model could account for the tissue-specific patterns of segregation bias observed in mouse models (Table 1). These data together display some structuring across tissues and mtDNA pairings. If haplotypes are classified according to which proliferates in liver (the most consistently observed segregation), three classes of behaviour are broadly observed across tissues. Some (kidney, testis) consistently follow the liver segregation pattern regardless of haplotype pairing. Others (spleen, blood, pancreas) show a different direction to liver in pairings involving NZB, and follow liver in other pairings. Still other tissues (muscle, brain, heart) show a different direction to liver in pairings involving JF1, RR and HB, and are comparatively neutral otherwise.

These tissue classes are broadly separable by cellular, and mtDNA, turnover rates, as found in (12). Pursuing this link, we ranked tissues by their estimated selectivity for bioenergetic capacity, a combination of cell turnover (allowing selection) and bioenergetic demand (imposing selection) – see

**Methods.** We then found that the predicted saddle structure from the RTS model was reflected in the observed mouse data (Figure 2B). We assigned haplotype pairs to one of three classes (mtDNA 1 more selfish, similar to, or less selfish than mtDNA type 2) and found that 74% (37/50) of nonzero segregation bias observations were accounted for by this classification. The overall classification success rate (0.66, 61/92) was significantly better than random assignment ( $P = 2.5 \times 10^{-5}$ , binomial test with  $B(92, 0.504)$  reflecting the proportion of nonzero observations). This fit to observed data predicts relative selfishness differences between pairs of observed haplotypes (i.e., the ordering on the horizontal axis of Figure 2B): (JF1, RR, HB, ST) < C57; (BG, LE, NZW)  $\simeq$  C57; (BALB, 129S6) < NZB.

### G-quadruplexes in the mtDNA control region influence replication-transcription balance

We next asked what specific sequence features might be associated with  $\lambda$ , the selfishness of a given mtDNA molecule, and why. In particular, we considered sequence features that may affect the balance of the replication-transcription switch mentioned above (26). Recent discussion has focussed on a particular section of the mtDNA control region called CSB2 (conserved sequence box II) (22,24–25) (Figure 3A), which has long been suggested to regulate the rate of human mtDNA synthesis (56).

A feature of CSB2 is its potential to form a hybrid G-quadruplex structure involving the complementary mtDNA strand and its nascent RNA transcript (56). G-quadruplexes are DNA secondary structures consisting of stacked planar guanine tetrads that play a number of regulatory roles, including in nuclear DNA replication, telomeres and transcription in a range of metazoans (57,58). CSB2 typically involves a series of cytosine bases broken by a central feature:  $C_6 - _9N_1 - _3C_5 - _6$ , with the corresponding guanine bases in the complementary strand and nascent transcript capable of forming a quadruplex structure. In mice, for example, the central feature usually consists of one or two adenine bases.

Previous studies have noted that differences in the number of guanine bases in CSB2 are associated with segregation bias (reversion) between mtDNA sequences in human stem cells (22,24–25). We hypothesise, cf. (56), that the stability of the associated G-quadruplex structure plays a role in determining the replication-transcription balance (selfishness) for a particular mtDNA sequence (Figure 3B).

*In vitro* experiments exploring transcription termination for a large variety of CSB2 structures found that termination occurred more readily for regions with more guanine bases (59). Independent *in vitro* mtDNA assays showed that shorter guanine tracts resulted in fourfold decreased replication primer formation (22). Biochemical observations confirm that G-quadruplexes formed with more guanine residues are more thermodynamically stable (60), and sequence-independent investigation shows that smaller nucleotide loops between guanine residues also stabilise the G-quadruplex (61). Taken together, these observations suggest a picture where more guanine residues, more tightly associated, produce stronger G-quadruplex structures (62), which more readily stall transcription (59) and thus more readily

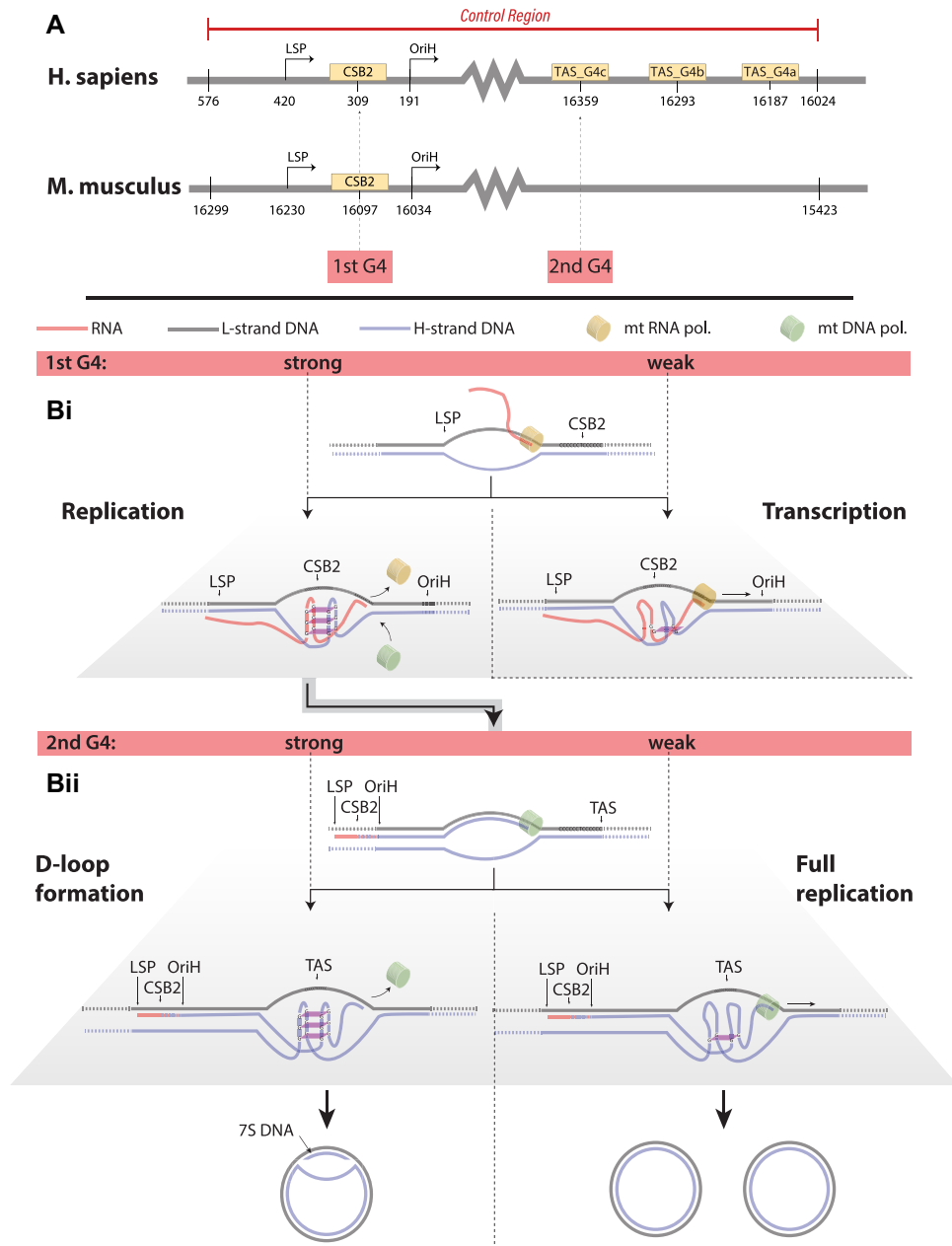
yield replication primers (22) (Figure 3B). This stalling may result from physical factors including tension in the elongating strand, or torsion or conformational changes induced by the hybrid G-quadruplex, either of which effects would be more potent with a ‘stronger’, less easily resolvable G-quadruplex. Hence, sequence features that favour transcription termination naturally make an mtDNA molecule more ‘selfish’: both limiting gene expression and promoting replication (Figure 3B).

CSB2 has received the most attention in the literature, but our hypothesised mechanism also predicts a potential role for other G-quadruplex forming mtDNA regions. Once a primer has formed after stalling at CSB2, and DNA replication has begun, any further unresolved stalling elsewhere will disfavour replication (Figure 3B). We therefore predict that a strong G-quadruplex at CSB2 will increase selfishness, but that strong G-quadruplexes elsewhere will decrease selfishness by challenging DNA replication. In particular,  $C_{3+}N_{1-7}C_{3+}$  potential intermolecular G-quadruplex forming motifs are found near the termination associated sequence (TAS) region in mtDNA also (see below). Our model predicts that weaker G-quadruplexes in this region will increase selfishness.

### G-quadruplex sequences in the control region predict segregation patterns in mice and human stem cells

To investigate evidence pertaining to our hypothesised mechanism, we first analysed the mtDNA sequences of mouse haplotypes from the pairings in Figure 2B. Alignment of the mouse mtDNA haplotypes with Clustal Omega (41) found diversity in CSB2 but not in any other potential G-quadruplex forming parts of the control region (Figure 3A). The fit to our RTS model predicted selfishness relationships (JF1, RR, HB, ST) < C57; (BG, LE, NZW)  $\simeq$  C57; (BALB, 129S6) < NZB. Correspondingly, relative to C57, JF1, RR and ST all have a destabilising additional adenine base in the centre of their CSB2 G-quadruplex region (RR has an additional proximal deletion of a cytosine base), and HB has a large insertion proximal to CSB2 that may influence G-quadruplex formation (Figure 4A). BG, LE, and NZW had no changes in the CSB2 region relative to C57. We thus found that 6/9 of fitted selfishness differences could be explained by sequence features decreasing the stability of the CSB2 G-quadruplex (59–61) (7/9 if the HB insertion is included; Figure 4A). We were unable to find predictive CSB2 differences between BALB, 129S6, and NZB, but in the particular case of NZB, the previously characterised functional difference between NZB and BALB may provide an alternative target for higher-level selection (63).

To further test our hypothesised link between G-quadruplex stability and selfishness, we analysed data from recent experiments in human embryonic stem cells (hESC) where segregation bias is observed. We reconstructed and aligned the haplotypes involved (see Methods), finding sequence diversity in both in the CSB2 and other potential G-quadruplex forming parts of the control region, specifically in the TAS regions, which we label TAS.G4a-c (Figure 3A). Refs. (22,24,25) have noted that a higher number of CSB2 guanine bases in these pairings is associated with increased propensity for segregation bias. Following our pro-

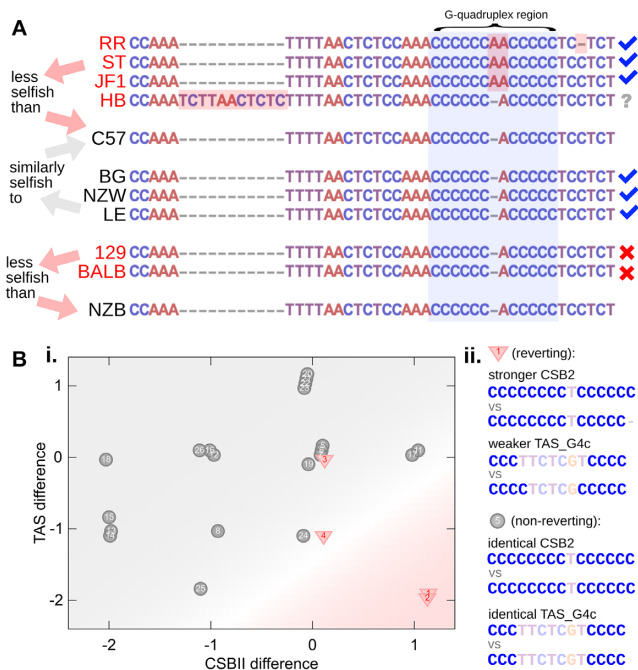


**Figure 3.** Control region features and hypothesised association with ‘selfishness’. **(A)** Selected positions in the human and mouse mtDNA control regions that possess sequence motifs expected to facilitate hybrid G-quadruplex formation. **(B)** G-quadruplex influence on replication-transcription balance. **(i)** After transcription initiation, a strong G-quadruplex at CSB2 may lead to RNA polymerase stalling and termination of transcription, while a weaker G-quadruplex allows transcription to proceed. Aborted transcription leads to primer formation, initiating mtDNA replication. **(ii)** After replication is initiated, further strong G-quadruplex regions may similarly interfere with DNA polymerase progress, causing D-loop formation rather than full replication. A single, strong G-quadruplex at CSB2 may thus be viewed as the strongest facilitator of replication.

posed mechanism, we asked whether taking the TAS region into account can improve the predictions of reversion. Of course, an absence of observed reversion in these experiments does not necessarily imply no segregation bias, because the starting amounts of maternal mtDNA may be small enough for random drift to remove even in the absence of a selective advantage. Further, the time scale of hESC culture is likely too short for the sort of tissue level segregation observed in mice to occur.

We used logistic regression with the number of guanine bases in CSB2 and the potential G-quadruplex regions in TAS (Figure 3A) as predictors of reversion. The model including CSB2 and TAS\_G4c regions had the lowest Bayesian Information Criterion, suggesting the strongest statistical support (Supplementary Table S1). We found that the CSB2-alone model had a pseudo  $R^2$  value of 0.26, but when TAS\_G4c was included this rose to 0.47. The CSB2-TAS\_G4c model accounted for 92% (24/26) observed re-





**Figure 4.** G-quadruplex features predict segregation patterns in mouse and human. (A) CSB2 and surrounding loci in mouse haplotypes from Table 1. Predicted selfishness relationships from fitting tissue-specific observations to the RTS model are given on the left. Candidate features modulating G-quadruplex stability are highlighted in red. Ticks, crosses, and question mark denote whether fitted selfishness is predicted by CSB2 region features. (B) i. Sequence features in pairs of mtDNA in human cell lines which are observed to show reversion via segregation bias (triangles) or not (circles). Labels denote specific pairings (see Methods). Horizontal axis gives difference in guanine count in CSB2 G-quadruplex region; vertical axis gives difference in guanine count in TAS G-quadruplex region. Shaded background shows the best-fit logistic regression model (see text) predicting reversion (pink) or no reversion (grey). ii. Example potential G-quadruplex sequences in a reverting and non-reverting case.

version behaviours (nonzero and zero), performing significantly better than random chance ( $P = 0.02$ , binomial test with  $B(26, 0.74)$ , reflecting the proportion of nonzero observations), and than the CSB2-alone model. When trained on  $10^3$  samples of half the observed haplotype pairs from (22) (half the reverting and half the non-reverting pairings), the CSB2-TAS had a mean prediction accuracy of 0.93 for reversion in the independent test pairings (Figure 4Bi). Together, these results support our hypothesis that a strong G-quadruplex in CSB2, and weak G-quadruplexes elsewhere, assign a proliferative advantage to an mtDNA sequence (Figure 4Bii).

Finally, we considered other species where segregation bias has been investigated. These include pig (19) and cattle (18) systems. In pig, some suggestion of tissue-specific segregation bias was observed but we were unable to show that it passed a significance threshold. Correspondingly, we found no diversity in candidate G-quadruplex sections in CSB2 for the haplotypes involved, although a polymorphism in TAS\_G4c was present that could potentially lead to smaller selfishness differences (Supplementary Figure S7A). By contrast, in cattle, segregation bias was observed in embryonic development, favouring *Bos taurus* mtDNA

over *Bos indicus*, and we identified a sequence difference in a candidate G-quadruplex region predicting a corresponding selfishness difference between mitotypes (Supplementary Figure S7B).

## DISCUSSION

We have proposed a mechanism whereby the replication-transcription balance ('selfishness') of an mtDNA molecule can affect its relative proliferation in heteroplasmic cells. The direction and magnitude of this effect in tissues is modulated by cellular selective pressure for bioenergetic capacity, which depends on cell turnover and bioenergetic demand. More selfish genomes proliferate in the absence of cell-level pressure; less selfish genomes, which support cellular function, proliferate when cell-level pressure is higher. Together, these effects can explain much of the diverse tissue-specific behaviour observed in mouse models.

We hypothesise, following biochemical arguments and observations in stem cells, that sequence features affecting the stability of G-quadruplexes in the mtDNA control region play a role in determining the replication-transcription balance of a given mtDNA sequence. These features—particularly, the number of guanine bases in the CSB2 and TAS regions—predict selfishness in mouse haplotypes and reversion propensity in hESCs.

It is likely that segregation bias depends on multiple sequence features as well as nuclear factors and metabolic context (63). Indeed, several nuclear-encoded genes have been found to influence the dynamics of heteroplasmic populations in somatic tissue (9,14–15) and the germline (64). The organelle- and cell-level selection we model are two nuclear-mediated strategies to mitigate against selfish mtDNA replication (31,32), and as such exist in the broader framing of co-operation and conflict between mtDNA and nuclear genes (recently reviewed in (32,65)).

Our model only suggests that differences in replication-transcription balance may account for some of these observed behaviours. We do not intend to claim that G-quadruplex stability is the only or even the leading factor determining segregation bias, but have shown that several observed features of segregation behaviour can be explained by these features, and that a plausible biophysical mechanism may underlie this predictive power. It is well known that mtDNA variants compromising functionality experience selective pressure (for example, the 3243 mutation in humans (66,67)). When one sequence has variants that directly challenge bioenergetic function or other aspects of cellular fitness, these may be acted upon by selective mechanisms and provide contributions to segregation behaviour in addition to those from our model. However, as we describe above, we expect our general findings (concerning the relative behaviour of two mtDNA types) to remain robust to global changes in cellular behaviour that are not specifically linked to only one mtDNA type. For example, a global reduction in mtDNA transcription (perhaps through limited polymerase expression) will not affect the relative behavioural differences between the two mtDNA types from which our predicted behaviour emerges.

Further, other potential G-quadruplex-forming regions also exist throughout human and other mtDNA sequences

(62,68); while our hypothesis focusses on the control region, the relative stabilities of these features across diverse genotypes may also play a role in mtDNA population dynamics, proposing further hypotheses for pursuit as more data becomes available. To our knowledge, no software tools yet exist that can make thermodynamic predictions about intermolecular DNA-RNA quadruplex stability (although a predictor for such structures does exist (69)). However, as we have argued above, insights from the more general literature on intramolecular G-quadruplexes (60–62) are informative, entropic arguments suggest that long G-tracts increase stability, and this theory is borne out by mtDNA-specific experimental observations of subsequent replication behaviour (22).

Our results are compatible with experiments in yeast (37) and *Drosophila* (21) exploring the tension between selfish drive and higher-level selection in shaping mtDNA populations. Interestingly, recent work exploring human mtDNA germline selection (70) observed pronounced differences between mother and child allele frequencies of several loci associated with potential G-quadruplex forming parts of the control region. While we did not find polymorphisms affecting the CSB2 G-quadruplex in the set of transmitted features in (70), there are several instances of transmission differences for a polymorphism (T16362C) in the section we label TAS\_G4c, and in other sections bearing potential G-quadruplex sequence motifs. These observations are compatible with the idea that stability of control region G-quadruplexes influence relative mtDNA proliferation, and suggest further candidate regions for investigation in addition to those we consider here.

Accordingly, our model makes several testable predictions for future experiments: most generally, that mtDNA sequences with more stable CSB2 G-quadruplexes, and less stable G-quadruplexes elsewhere, will show the tissue-specific segregation behaviour associated with less selfish molecules (proliferation in highly selective tissues, loss in weakly selective tissues).

The question of matching mtDNA sequences ('haplotype matching') is frequently discussed in the application of mtDNA gene therapies (71). Haplotype matching is proposed as an attempt to avoid problematic segregation bias in gene therapies by selecting a 'donor' mtDNA type for which maternal reversion through segregation bias is unlikely. Our model agrees with recent discussion on this topic (24,25) suggesting that the number of guanine bases in CSB2, rather than general sequence similarity, may be a promising consideration for minimising reversion. We would also suggest consideration of the TAS region that our work indicates may also affect segregation bias. Substantial diversity exists in these regions in human mtDNA accessions in NCBI (Supplementary Figure S6) which partially correlates with overall mtDNA haplotype.

While this manuscript was in press, two exciting related developments appeared in the scientific literature. The first reports an mtDNA sequence feature in transmissible canine cancer that seems likely to confer a replicative advantage: an insertion of two C bases at position 16660, next to an existing C tract in the control region (72). It may be possible that this CC insertion may influence the formation of a hybrid G-quadruplex structure at this site, suggesting a pos-

sible link to our proposed mechanism modulating mtDNA replication-transcription balance. The second is the report of precise gene editing in mtDNA (73). This groundbreaking development will allow the design and investigation of synthetic mtDNA sequences to explore the mechanisms of segregation bias on an unprecedented scale, providing a clear future route to testing our hypothesised mechanism.

Our RTS model adds an explicit description of mtDNA gene expression to a core of stochastic models for functional (37,50–54) and genetic (36,47–48,74–77) dynamics of mixed mitochondrial populations. There is a growing intersection between this body of work and the wide literature on multilevel selection on competing replicating agents in an evolutionary context (78,79). Theoretical work on model unicellular organisms has explored the effect of mitochondrial bottlenecks, paternal transmission, and cellular selection (77). Previous work in plants, where mtDNA recombination produces diverse cellular mtDNA admixtures of potentially different selfishnesses, has shown that multilevel intra- or intercellular selection provides a way of combating selfish proliferation (36). The pronounced role for mitochondrial dynamics in imposing intra-cellular quality control (30), captured by many of these models (50–54,75), suggests that tissue-specific mitochondrial dynamics may play an important role in this multilevel selection and mitigation of selfish proliferation (32,80).

## DATA AVAILABILITY

Simulation code for the organelle- and cell-based RTS models, code for bioinformatic and statistical analysis of data from Ref. (25) and the NCBI, visualisation code, and mouse and stem cells alignments are freely available at <https://github.com/StochasticBiology/selfish-mtdna>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank Paul Una for valuable contributions.

## FUNDING

European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme [Grant agreement No. 805046 (EvoConBio) to IGJ]. Funding for open access charge: ERC [805046].

*Conflict of interest statement.* None declared.

## REFERENCES

1. Johnston,I.G. and Burgstaller,J.P. (2019) Evolving mtDNA populations within cells. *Biochem. Soc. Trans.*, **47**, 1367–1382.
2. Stewart,J.B. and Chinnery,P.F. (2015) The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat. Rev. Genet.*, **16**, 530.
3. Wallace,D.C. and Chalkia,D. (2013) Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Csh. Perspect. Biol.*, **5**, a021220.
4. Burgstaller,J.P., Johnston,I.G. and Poulton,J. (2014) Mitochondrial DNA disease and developmental implications for reproductive strategies. *Mol. Hum. Reprod.*, **21**, 11–22.

5. Jenuth, J.P., Peterson, A.C. and Shoubridge, E.A. (1997) Tissue-specific selection for different mtDNA genotypes in heteroplasmic mice. *Nat. Genet.*, **16**, 93.
6. Acton, B., Lai, I., Shang, X., Jurisicova, A. and Casper, R. (2007) Neutral mitochondrial heteroplasmy alters physiological function in mice. *Biol. Reprod.*, **77**, 569–576.
7. Sharpley, M.S., Marciniak, C., Eckel-Mahan, K., McManus, M., Crimi, M., Waymire, K., Lin, C.S., Masubuchi, S., Friend, N., Koike, M. et al. (2012) Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and cognition. *Cell*, **151**, 333–343.
8. Takeda, K., Takahashi, S., Onishi, A., Hanada, H. and Imai, H. (2000) Replicative advantage and tissue-specific segregation of RR mitochondrial DNA between C57BL/6 and RR heteroplasmic mice. *Genetics*, **155**, 777–783.
9. Battersby, B.J., Loredó-Osti, J. and Shoubridge, E.A. (2003) Nuclear genetic control of mitochondrial DNA segregation. *Nat. Genet.*, **33**, 183–186.
10. Battersby, B.J., Redpath, M.E. and Shoubridge, E.A. (2005) Mitochondrial DNA segregation in hematopoietic lineages does not depend on MHC presentation of mitochondrially encoded peptides. *Hum. Mol. Genet.*, **14**, 2587–2594.
11. Inoue, K., Ogonuki, N., Yamamoto, Y., Takano, K., Miki, H., Mochida, K. and Ogura, A. (2004) Tissue-specific distribution of donor mitochondrial DNA in cloned mice produced by somatic cell nuclear transfer. *Genesis*, **39**, 79–83.
12. Burgstaller, J.P., Johnston, I.G., Jones, N.S., Albrechtova, J., Kolbe, T., Vogl, C., Futschik, A., Mayrhofer, C., Klein, D., Sabitzer, S. et al. (2014) MtDNA segregation in heteroplasmic tissues is common in vivo and modulated by haplotype differences and developmental stage. *Cell Rep.*, **7**, 2031–2041.
13. Pan, J., Wang, L., Lu, C., Zhu, Y., Min, Z., Dong, X. and Sha, H. (2019) Matching mitochondrial DNA haplotypes for circumventing tissue-specific segregation bias. *Science*, **13**, 371–379.
14. Jokinen, R., Marttinen, P., Sandell, H.K., Manninen, T., Teerenhovi, H., Wai, T., Teoli, D., Loredó-Osti, J.C., Shoubridge, E.A. and Battersby, B.J. (2010) Gimap3 regulates tissue-specific mitochondrial DNA segregation. *PLoS Genet.*, **6**, e1001161.
15. Jokinen, R. and Battersby, B.J. (2013) Insight into mammalian mitochondrial DNA segregation. *Ann. Med.*, **45**, 149–155.
16. de Stordeur, E., Solignac, M., Monnerot, M. and Mounolou, J.-C. (1989) The generation of transplasmic *Drosophila simulans* by cytoplasmic injection: Effects of segregation and selection on the perpetuation of mitochondrial DNA heteroplasmy. *Mol. Gen. Genet. MGG*, **220**, 127–132.
17. Solignac, M., Génemont, J., Monnerot, M. and Mounolou, J.-C. (1987) *Drosophila* mitochondrial genetics: evolution of heteroplasmy through germ line cell divisions. *Genetics*, **117**, 687–696.
18. Ferreira, C.R., Burgstaller, J.P., Perecin, F., Garcia, J.M., Chiaratti, M.R., Méo, S.C., Müller, M., Smith, L.C., Meirelles, F.V. and Steinborn, R. (2010) Pronuclear segregation of donor mitochondria introduced by bovine ooplasmic transfer to the female germ-line. *Biol. Reprod.*, **82**, 563–571.
19. Takeda, K., Tasai, M., Iwamoto, M., Akita, T., Tagami, T., Nirasawa, K., Hanada, H. and Onishi, A. (2006) Transmission of mitochondrial DNA in pigs and progeny derived from nuclear transfer of Meishan pig fibroblast cells. *Mol. Reprod. Dev.: Incorp. Gamete Res.*, **73**, 306–312.
20. Cagnone, G., Tsai, T.-S., Srirattana, K., Rossello, F., Powell, D.R., Rohrer, G., Cree, L., Trounce, I.A. and John, J.C.S. (2016) Segregation of naturally occurring mitochondrial DNA variants in a mini-pig model. *Genetics*, **202**, 931–944.
21. Ma, H. and O'Farrell, P.H. (2016) Selfish drive can trump function when animal mitochondrial genomes compete. *Nat. Genet.*, **48**, 798.
22. Kang, E., Wu, J., Gutierrez, N.M., Koski, A., Tippner-Hedges, R., Agaronyan, K., Platero-Luengo, A., Martinez-Redondo, P., Ma, H., Lee, Y. et al. (2016) Mitochondrial replacement in human oocytes carrying pathogenic mitochondrial DNA mutations. *Nature*, **540**, 270.
23. Hyslop, L.A., Blakeley, P., Craven, L., Richardson, J., Fogarty, N.M., Fragouli, E., Lamb, M., Wamaitha, S.E., Prathalingam, N., Zhang, Q. et al. (2016) Towards clinical application of pronuclear transfer to prevent mitochondrial DNA disease. *Nature*, **534**, 383.
24. Hudson, G., Takeda, Y. and Herbert, M. (2019) Reversion after replacement of mitochondrial DNA. *Nature*, **574**, E8–E11.
25. Kang, E., Koski, A., Amato, P., Temiakov, D. and Mitalipov, S. (2019) Reply to: Reversion after replacement of mitochondrial DNA. *Nature*, **574**, E12–E13.
26. Agaronyan, K., Morozov, Y.I., Anikin, M. and Temiakov, D. (2015) Replication-transcription switch in human mitochondria. *Science*, **347**, 548–551.
27. Falkenberg, M. (2018) Mitochondrial DNA replication in mammalian cells: overview of the pathway. *Essays Biochem.*, **62**, 287–296.
28. Farge, G., Mehmedovic, M., Baclayon, M., van den Wildenberg, S.M., Roos, W.H., Gustafsson, C.M., Wuite, G.J. and Falkenberg, M. (2014) In vitro-reconstituted nucleoids can block mitochondrial DNA replication and transcription. *Cell Rep.*, **8**, 66–74.
29. Kühl, I., Miranda, M., Posse, V., Milenkovic, D., Mourier, A., Siira, S.J., Bonkamp, N.A., Neumann, U., Filipovska, A., Polosa, P.L. et al. (2016) POLRMT regulates the switch between replication primer formation and gene expression of mammalian mtDNA. *Science Advances*, **2**, e1600963.
30. Twig, G., Hyde, B. and Shirihai, O.S. (2008) Mitochondrial fusion, fission and autophagy as a quality control axis: the bioenergetic view. *BBA-Bioenergetics*, **1777**, 1092–1097.
31. Haig, D. (2016) Intracellular evolution of mitochondrial DNA (mtDNA) and the tragedy of the cytoplasmic commons. *BioEssays*, **38**, 549–555.
32. Havird, J.C., Forsythe, E.S., Williams, A.M., Werren, J.H., Dowling, D.K. and Sloan, D.B. (2019) Selfish mitonuclear conflict. *Curr. Biol.*, **29**, R496–R511.
33. Werren, J.H., Nur, U. and Wu, C.-I. (1988) Selfish genetic elements. *Trends Ecol. Evol.*, **3**, 297–302.
34. Hurst, L. (1995) Selfish genetic elements and their role in evolution: the evolution of sex and some of what that entails. *Philos. T. Roy. Soc. B*, **349**, 321–332.
35. Schnable, P.S. and Wise, R.P. (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends Plant. Sci.*, **3**, 175–180.
36. Albert, B., Godelle, B., Atlan, A., De Paepe, R. and Gouyon, P.H. (1996) Dynamics of plant mitochondrial genome: model of a three-level selection process. *Genetics*, **144**, 369–382.
37. Taylor, D.R., Zeyl, C. and Cooke, E. (2002) Conflicting levels of selection in the accumulation of mitochondrial defects in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 3690–3694.
38. Poulton, J., Deadman, M., Ramacharan, S. and Gardiner, R. (1991) Germ-line deletions of mtDNA in mitochondrial myopathy. *Am. J. Hum. Genet.*, **48**, 649.
39. Gitschlag, B.L., Kirby, C.S., Samuels, D.C., Gangula, R.D., Mallal, S.A. and Patel, M.R. (2016) Homeostatic responses regulate selfish mitochondrial genome dynamics in *C. elegans*. *Cell Metab.*, **24**, 91–103.
40. Kang, E., Wu, J., Gutierrez, N.M., Koski, A., Tippner-Hedges, R., Agaronyan, K., Platero-Luengo, A., Martinez-Redondo, P., Ma, H. and Lee, Y. et al. (2019) Author correction: Mitochondrial replacement in human oocytes carrying pathogenic mitochondrial DNA mutations. *Nature*, **567**, E5–E9.
41. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M. and Söding, J. et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol. Syst. Biol.*, **7**, 539–545.
42. Wilkens, V., Kohl, W. and Busch, K. (2013) Restricted diffusion of OXPHOS complexes in dynamic mitochondria delays their exchange between cristae and engenders a transitory mosaic distribution. *J. Cell Sci.*, **126**, 103–116.
43. Hoitzing, H., Johnston, I.G. and Jones, N.S. (2015) What is the function of mitochondrial networks? A theoretical assessment of hypotheses and proposal for future research. *Bioessays*, **37**, 687–700.
44. Gelman, A. and Su, Y.-S. (2018) In: *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models*. R package version 1.10-1.
45. R Core Team (2015) In: *R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing*. Vienna.
46. Jackman, S. (2017) In: *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory United States Studies Centre*. University of Sydney Sydney, New South Wales, R package version 1.5.2.

47. Chinnery, P.F. and Samuels, D.C. (1999) Relaxed replication of mtDNA: a model with implications for the expression of disease. *Am. J. Hum. Genet.*, **64**, 1158–1165.
48. Johnston, I.G. and Jones, N.S. (2016) Evolution of cell-to-cell variability in stochastic, controlled, heteroplasmic mtDNA populations. *Am. J. Hum. Genet.*, **99**, 1150–1162.
49. Hoitzing, H., Johnston, I.G. and Jones, N.S. (2017) Stochastic models for evolving cellular populations of mitochondria: disease, development, and ageing. In: *Stochastic Processes, Multiscale Modeling, and Numerical Methods for Computational Cellular Biology*. Springer, pp. 287–314.
50. Mouli, P.K., Twig, G. and Shirihai, O.S. (2009) Frequency and selectivity of mitochondrial fusion are key to its quality maintenance function. *Biophys. J.*, **96**, 3509–3518.
51. Poovathingal, S.K., Gruber, J., Halliwell, B. and Gunawan, R. (2009) Stochastic drift in mitochondrial DNA point mutations: a novel perspective ex silico. *PLoS Comput. Biol.*, **5**, e1000572.
52. Tam, Z.Y., Gruber, J., Halliwell, B. and Gunawan, R. (2013) Mathematical modeling of the role of mitochondrial fusion and fission in mitochondrial DNA maintenance. *PloS one*, **8**, e76230.
53. Tam, Z.Y., Gruber, J., Halliwell, B. and Gunawan, R. (2015) Context-dependent role of mitochondrial fusion-fission in clonal expansion of mtDNA mutations. *PLoS Comput. Biol.*, **11**, e1004183.
54. Figge, M.T., Reichert, A.S., Meyer-Hermann, M. and Osiewacz, H.D. (2012) Deceleration of fusion–fission cycles improves mitochondrial quality control during aging. *PLoS Comput. Biol.*, **8**, e1002576.
55. Burgstaller, J.P., Kolbe, T., Havlicek, V., Hembach, S., Poulton, J., Piálek, J., Steinborn, R., Rüllicke, T., Brem, G., Jones, N.S. and et al. (2018) Large-scale genetic analysis reveals mammalian mtDNA heteroplasmy dynamics and variance increase through lifetimes and generations. *Nat. Commun.*, **9**, 2488.
56. Wanrooij, P.H., Uhler, J.P., Shi, Y., Westerlund, F., Falkenberg, M. and Gustafsson, C.M. (2012) A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucleic Acids Res.*, **40**, 10334–10344.
57. Rhodes, D. and Lipps, H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627–8637.
58. Bryan, T.M. (2019) Mechanisms of DNA replication and repair: insights from the study of G-quadruplexes. *Molecules*, **24**, 3439.
59. Tan, B.G., Wellesley, F.C., Savery, N.J. and Szczelkun, M.D. (2016) Length heterogeneity at conserved sequence block 2 in human mitochondrial DNA acts as a rheostat for RNA polymerase POLRMT activity. *Nucleic Acids Res.*, **44**, 7817–7829.
60. Petraccone, L., Erra, E., Duro, I., Esposito, V., Randazzo, A., Mayol, L., Mattia, C.A., Barone, G. and Giancola, C. (2005) Relative stability of quadruplexes containing different number of G-tetrads. *Nucleosides Nucleotides Nucleic Acids*, **24**, 757–760.
61. Bugaut, A. and Balasubramanian, S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry*, **47**, 689–697.
62. Kikin, O., D’Antonio, L. and Bagga, P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.*, **34**, W676–W682.
63. Moreno-Loshuertos, R., Acín-Pérez, R., Fernández-Silva, P., Movilla, N., Pérez-Martos, A., de Córdoba, S.R., Gallardo, M.E. and Enriquez, J.A. (2006) Differences in reactive oxygen species production explain the phenotypes associated with common mouse mitochondrial DNA variants. *Nat. Genet.*, **38**, 1261.
64. Latorre-Pellicer, A., Lechuga-Vieco, A.V., Johnston, I.G., Hämäläinen, R.H., Pellico, J., Justo-Méndez, R., Fernández-Toro, J.M., Clavería, C., Guaras, A., Sierra, R. et al. (2019) Regulation of Mother-to-Offspring transmission of mtDNA heteroplasmy. *Cell Metab.*, **30**, 1120–1130.
65. Rand, D.M. and Mossman, J.A. (2020) Mitonuclear conflict and cooperation govern the integration of genotypes, phenotypes and environments. *Philos. Trans. R. Soc. B*, **375**, 20190188.
66. Pyle, A., Taylor, R.W., Durham, S.E., Deschauer, M., Schaefer, A.M., Samuels, D.C. and Chinnery, P.F. (2007) Depletion of mitochondrial DNA in leucocytes harbouring the 3243A G mtDNA mutation. *J. Med. Genet.*, **44**, 69–74.
67. Rajasimha, H.K., Chinnery, P.F. and Samuels, D.C. (2008) Selection against pathogenic mtDNA mutations in a stem cell population leads to the loss of the 3243A G mutation in blood. *Am. J. Hum. Genet.*, **82**, 333–343.
68. Brázda, V., Kolomazník, J., Lýsek, J., Bartas, M., Fojta, M., Št’astný, J. and Mergny, J.-L. (2019) G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics*, **35**, 3493–3495.
69. Kudlicki, A.S. (2016) G-quadruplexes involving both strands of genomic DNA are highly abundant and colocalize with functional sites in the human genome. *PLoS One*, **11**, e0146174.
70. Wei, W., Tuna, S., Keogh, M.J., Smith, K.R., Aitman, T.J., Beales, P.L., Bennett, D.L., Gale, D.P., Bitner-Glindzicz, M.A., Black, G.C. et al. (2019) Germline selection shapes human mitochondrial DNA diversity. *Science*, **364**, eaau6520.
71. Røyrvik, E., Burgstaller, J.P. and Johnston, I.G. (2016) mtDNA diversity in human populations highlights the merit of haplotype matching in gene therapies. *MHR: Basic Sci. Reprod. Med.*, **22**, 809–817.
72. Strakova, A., Nicholls, T.J., Baez-Ortega, A., Leathlobhair, M.N., Sampson, A.T., Hughes, K., Bolton, I.A., Gori, K., Wang, J. and Airikkala-Otter, I. et al. (2020) Recurrent horizontal transfer identifies mitochondrial positive selection in a transmissible cancer. *Nature communications*, **11**, 1–11.
73. Mok, B.Y., de Moraes, M.H., Zeng, J., Bosch, D.E., Kotrys, A.V., Raguram, A., Hsu, F., Radey, M.C., Peterson, S.B. and Mootha, V.K. et al. (2020) Abacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature*, 1–7.
74. Johnston, I.G., Burgstaller, J.P., Havlicek, V., Kolbe, T., Rüllicke, T., Brem, G., Poulton, J. and Jones, N.S. (2015) Stochastic modelling, Bayesian inference, and new in vivo measurements elucidate the debated mtDNA bottleneck mechanism. *Elife*, **4**, e07464.
75. Aryaman, J., Bowles, C., Jones, N.S. and Johnston, I.G. (2019) Mitochondrial network state scales mtDNA genetic dynamics. *Genetics*, **212**, 1429–1443.
76. Atlan, A. and Couvet, D. (1993) A model simulating the dynamics of plant mitochondrial genomes. *Genetics*, **135**, 213–222.
77. Roze, D., Rousset, F. and Michalakis, Y. (2005) Germline bottlenecks, biparental inheritance and selection on mitochondrial variants: a two-level selection model. *Genetics*, **170**, 1385–1399.
78. Szathmáry, E. and Demeter, L. (1987) Group selection of early replicators and the origin of life. *J. Theor. Biol.*, **128**, 463–486.
79. Szathmáry, E. (2015) Toward major evolutionary transitions theory 2.0. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 10104–10111.
80. Johnston, I.G. (2019) Tension and resolution: dynamic, evolving populations of organelle genomes within plant cells. *Mol. Plant*, **12**, 764–783.